УДК 004.03+004.056

Кульчицький Б. В.
Куперштейн Л. М.

ДО ПРОБЛЕМИ ФОРМУВАННЯ НАБОРУ ДАНИХ ДЛЯ ДОСЛІДЖЕННЯ DDOS-АТАК

Вінницький національний технічний університет

Анотація

В роботі розглянуто підходи щодо перевірки запропонованих методів виявлення атак. Проаналізовано наявні набори даних, які використовуються для створення систем виявлення DDoS-атак. Також, проаналізовано декілька інструментів, що використовуються для реалізації чи моделювання DDoS-атак для збору даних.

Ключові слова: DDoS-атака, набір даних, система виявлення вторгнень, генерування трафіку.

Abstract

The paper considers approaches to checking the proposed method of detecting attacks. The existing datasets that scientists use to create DDoS-attack detection systems are analyzed. Also, there are several tools used to implement or simulate DDoS-attacks for data collection.

Keywords: DDoS-attack, dataset, intrusion detection system, traffic generation.

Вступ

DDoS-атаки стали серйозною причиною проблем і загроз безпеці для підприємств, банків та бізнесу через Інтернет. Зловмисники постійно вдосконалюють свої навички, використовуючи покращенні прийоми атаки, щоб запустити такий величезний обсяг трафіку, який би зміг перемогти існуючі захисні рішення [1]. Атака має ряд наслідків: істотно впливає на ефективність сайту, знижується репутація організації та одне з головних це втрата доходу компанії.

Всякий раз, коли дослідник пропонує будь-який новий метод виявлення або захисту запропонований підхід повинен бути реалізований у вигляді експерименту. Для його перевірки та оцінки необхідні набори даних (dataset), які можуть бути попередньо зібрані чи згенеровані програмними засобами. Оскільки дуже складно і дорого провести справжню DDoS-атаку, більшість організацій і дослідників в результаті використовують методику моделювання, щоб імітувати реальну атаку. DDoS-атаки еволюціонують та стають більш складними для визначення системами виявлення вторгнень (IDS), через недостатність відомостей про актуальні атаки. Доступні та згенеровані набори трафіку стають все більш важливими при розробці нової IDS.

Метою даного дослідження є аналіз існуючих наборів даних та програмних засобів моделювання DDoS-атак для більш ефективного їх використання в задачах виявлення та попередження мережевих вторгнень, а також прийняття рішення щодо їх подальшого вдосконалення.

Результати дослідження

Система виявлення вторгнень (IDS) може бути програмним або апаратним забезпеченням для моніторингу та виявлення будь-якої мережевої загрози проти системи. Існують два основні підходи до виявлення: на основі сигнатури та на основі аномалій. Методика виявлення на основі сигнатури порівнює відомі відомості з вже захопленими сигнатурами, що зберігаються в базі даних. Ця методика здатна лише виявляти відомі атаки і має низький рівень помилкової тривоги. На відміну від першого підходу, система виявлення аномалій визначає будь-яку не типову для системи поведінку. Таким чином, вона здатна виявляти невідомі атаки, але з більш високою помилковою тривогою.

Виділяють чотири підходи перевірки запропонованого методу виявлення, що використовуються при розробці IDS [2]:

– Математичні моделі які носять теоретичний характер. У таких моделях дана система, додатки, платформи та умови, які моделюються символічно і потім перевіряються математично.

– Моделювання забезпечує відтворювану і контрольовану структуру для експериментів на основі мережі на одній комп'ютерній системі. Експерименти на основі моделювання дуже прості у налаштуванні та керуванні. Це надає розробникам гнучкість у виконанні експериментів над прототипами та відкинути багато поганих альтернатив своєчасно, перш ніж здійснити повну реалізацію. При моделюванні використовуються моделі ключових функцій операційної системи, механізми ядра, віртуальні платформи та синтетичні умови для експериментів.

– Емуляція - це інтеграція моделювання та реальних систем. У емуляції реальні елементи операційної системи та реальні програми поєднуються з нереальними та імітаційними елементами, такими як м'які мережні посилання, віртуальні проміжні вузли та нереальний фоновий трафік. Моделювання виконується у віртуальному модельованому часі, тоді як емуляція працює в режимі реального часу. Техніка емуляції має на увазі фактор масштабованості, оскільки дуже складно розширити топологію комп'ютерних систем за певними межами і не можна зробити топологію такою ж великою, як постачальник Інтернет-послуг.

– Реальні системи забезпечують реалістичні умови мережі, реальні операційні системи, додатки та платформи, як результат найкраще підходять для мережевих експериментів. Однак існують певні обмеження використання реальних систем для експериментів: зміна топології мережі неможлива для нового експерименту; дуже небезпечно проводити експерименти в Інтернеті з черв'яками, вірусами тощо, тому що вони можуть легко вийти з налаштованої мережі експерименту та пошкодити компоненти мережі в реальному часі; DDoS-атаки засновані на затопленні, можуть призвести до погіршення мережевих зв’язків.

Оскільки математичні моделі носять теоретичний характер, вони не можуть бути використані для таких мережевих експериментів. Техніка моделювання не може виконувати реальні програми, оскільки вона може наблизити лише певні апаратні та програмні функції. Емуляція забезпечує зручний спосіб використання реальних апаратних засобів і додатків, але його функціональність обмежується кількістю вузлів, типами апаратних засобів та різними конфігураціями, управлінням і відтворюваністю експериментів. Реальні системи ідеально підходять для перевірки досліджень на основі мережі, але вони обмежені в їх функціональності через їх складний характер.

Для отримання набору даних DDoS-атак дослідники можуть: використовувати спеціалізовані програмні засоби, які базуються на перерахованих вище підходах, для отримання необхідної їм типу атаки, так і використання вже зібраних даних, які містять як реальний так і згенерований трафік. Дані можуть бути використані для формування нових систем виявлення вторгнень, які здатні передбачити різні типи DDoS-атак.

Існують загальнодоступні набори даних, які дослідники використовують для тестування своєї методики та ефективності алгоритмів. Деякі набори даних отримуються з реальних атак, тоді як інші - об’єднання імітаційних атак. Тим не менш, є необхідність відзначити, що статистичні дані, представлені в наборі даних, як правило, відрізняються від реального мережевого трафіку. Наведено порівняння різних наборів даних, що використовуються при моделюванні DDoS-атак представлено в табл. 1.

Таблиця 1 – Порівняння різних наборів даних, що використовуються для виявлення DDoS-атак

Ім’я dataset	**Дата**	**Реальна чи змодельована**	**Типи DDoS-атак**	**Особливості**	**Переваги/Недоліки**
EPA http dataset	29 серпня 1995	Реальний	HTTP Flooding	• 46,014 GET запитів, 1622 POST запитів, 107 HEAD запитів, 6 невірних запитів; • розмір – 4.4 MB	+ Малий набір даних; – Не можна визначити законні та незаконні HTTP-запити. – Малий набір даних може обмежити ступінь виявлення атаки.
FIFA World Cup Dataset	30 квітня – 26 липня 1998	Реальний	HTTP Flooding	• 1.35 мілярдів запитів; • розмір – 307 MB	+ Роздільна здатність мітки часу 1 секунда.
KDD’99 Cup Dataset [3]	28 жовтня 1999	Змодельований	Back, Land, Neptune, Pod, Smurf, Teardrop	• вихід розділений на 5 категорії DOS, Probe, R2L, U2R і звичайний; • містить 38 типів атак; • розмір – 743 MB	+ Легко отримати. + Багато доступних типів атак. – Сильно незбалансований набір даних з 80% трафіку атак.
DoS_80_ timeseries-20020629	29 червня 2002 – 30 жовтня 2003	Реальний	Reflection TCP атака без прапорів	• часова серія з 80 DoS-атак; • розмір – 783.1 MB	+ Найкоротший часовий ряд з 1 мілісекундою зернистості.

Продовження табл. 1

CAIDA DDoS Attack Dataset [4]	4 серпня 2007	Змодельований	UDP flood	• складаються з анонімних даних протягом однієї години; • розмір – 21 GB	+ Ефективна для обробки великих DDoS-атак вище 5 Гб. + Сліди можна прочитати на будь-якому програмному забезпеченні, читаючи tcpdump. – Нормальний трафік недоступний. – Не включає в себе payload пакети.
NSL-KDD Dataset [5]	2009	Змодельований	Back, Land, Neptune, Process table, Worm, Apache2	• безперервна тривалість; • дискретний протокол; • дискретне обслуговування; • розмір – 124 MB	
DARPA_2009_DDoS_attack-20091105	5 листопада 2009	Реальний	SYN Flood	• SYN floods націлені на одну IP-адресу (172.28.4.7); • атака має фоновий трафік; • DDoS-трафік з 100 окремих IP-адрес; • розмір – 1.01 GB	+ Складається з атак кількох реальних джерел, отже, можливо дізнатися вектори атаки. – Атака націлена на одну жертву, не визначає загальну міцність мережі.
ISCX dataset [6]	11 червня – 17 червня 2010	Змодельований	HTTP, SMTP, SSH, IMAP, FTP	• практична мережа і трафік; • різні сценарії вторгнення; • розмір – 84.46 GBi	
DoS_80-20110715 [7]	15 липня 2011	Змодельований	TCP SYN/ ACK	• складається лише з одної атаки; • розмір – 32.31 GB	+ Визначено 8 відомих помилкових спрацьовувань. – Містить багато однакових пакетів.
TUIDS [8]	2012	Змодельований	TCP SYN, UDP Flood, Fraggle Flood, Smurf Flood	• справжні IP-адреси; • розмір – 66.7 GB	+ Містить різні типи DDoS-атак
FRGP_NTP_Flow_Data_anon-20131201	1 грудня 2013 – 28 лютого 2014	Анонімізований	NTP reflection атаки	• 3-денний NTP в потоці Argus по 10 Гбіт/с; • зловмисники запускають атаки надсилаючи запити monlist, що містять підроблені IP-адреси для хостів NTP; • розмір – 726.7 GB	+ Великий набір даних, що містять вектори для вимірювання декількох типів атак, включаючи спуфінг.
FRGP_SSDP_Reflection_DDoS_Attack_Traffic-20140930	30 вересня 2014	Змодельований	SSDP reflection атаки	• атакуючий потік на 10 Гб/с; • атака ініціюється за допомогою відкриття UPnP/SSDP за допомогою підробленого джерела IP для уразливих хостів, що запускають SSDP; • розмір – 26 GB	
Mirai-Bscanning-20160601	1 червня 2016 – 30 березня 2017	Реальний	TCP SYN	• тільки Mirai ідентифікований TCP SYN на портах 23 і 2323; • розмір – 1.1 GB	+ Містять дані реальних атак, отже, здатні запобігти майбутнім таким атакам. – Зібрані тільки Mirai-атак обмежує дослідників від інших атак.

Надійність вибраного набору даних для перевірки підходу залишається відкритим питанням. Дуже важливо вибрати відповідний набір даних для перевірки будь-якої запропонованої методики виявлення DDoS-атаки. В захопленому мережевому трафіку повинно містити суміш нормального та атакуючого трафіку у відповідній пропорції, і не повинно бути зміщень до певного типу трафіку. Але, дуже важко забезпечити відповідну суміш нормального і атакуючого трафіку в реальному наборі експериментів, оскільки не існує відомої формули для правильного моделювання мережевого трафіку. У дослідженні DDoS-атак широко використовуються такі набори даних як CAIDA, DARPA і TUIDS інші набори даних використовуються рідко, тому що вони досить застарілі або обмежені для

використання. Так в популярному наборі даних KDD Cup 1999 дуже важко розрізняти трафік атаки з нормального трафіку, оскільки набір даних не належно позначений [9].

Також можна використати доступні програми для генерування трафіку та побудови власної мережі. Спосіб реалізації моделювання або емуляції набору даних залежить від вимог до генерації типу трафіку і мети виконання експерименту. Застосовуючи моделювання та емуляцію мережі, імітований трафік може бути близьким до реального трафіку DDoS-атаки, але проблеми полягають не тільки в витратах на придбання середовища емуляції чи змодельованої мережі, але й на додаткових професійних знаннях, необхідних для створення, розгортання та експлуатування комплектів або системи керування. Тому, дослідники обирають вже реалізовані програмні засоби для проведення DDoS-атак. Інструменти, що використовуються при моделюванні чи реалізації DDoS-атаки включають: Ddosflowgen, OMNET++, Shaft, Tribe Flood, Network (TFN), LOIC, Trininty v3, Knight, WinCap and JpCap. В табл. 2 наведено порівняння різних інструментів DDOS-атаки з відповідними можливостями.

Таблиця 2 – Інструменти моделювання DDoS-атак

Інструмент моделювання	**Протокол**	**Атака**	**Можливості**
Ddosflowgen [10]	UDP, TCP	UDP flood, TCP requests, Mirai scans	– Можливість визначення кількості атакуючих мереж і налаштування параметрів таких як: коефіцієнт посилення, вектори атаки та кількість джерел мережевих атак. – Створює набори синтетичних даних трафіку з N переглядів. – Може обробляти атаки понад 1 Тб/с
OMNET++ [11]	UDP, TCP, ICMP	Атаки транспортного ріння	– Керована форма веб-серверу. – Здатний до моделювання TCP/IP.
Shaft	ICMP, UDP, TCP	TCP flood, UDP flood, ICMP flood	– Обробники та агенти спілкуються через UDP. – Рандомізує вихідний порт і IP-адреси в пакетах. – Фіксований розмір пакета під час атаки. – Перемикає керування основними серверами і портами в режимі реального часу тим самим ускладнюючи засоби виявлення вторгнень.
Tribe Flood Network (TFN)	TCP, UDP, ICMP	TCP SYN, ICMP flood, smurf	– Використовується для зменшення пропускної здатності та ресурсів. – Використовує інтерфейс командного рядка для атакуючого.
LOIC [12]	TCP, UDP, HTTP	UDP, TCP, HTTP flood	– Інструмент анонімної атаки на основі IRC. – Існує як двійкова, так і веб-версія
Trinity v3	UDP, TCP	TCP fragment, established and random flag floods, RST packet floods,	– TCP flood зроблені шляхом рандомізації всіх 32-бітів IP-адреси джерела. – Пакети Flood генерується за допомогою випадкових прапорів управління.
Knight	TCP, UDP	PUSH and flood TCP, SYN, UDP flooding	– Використовує Back Orifice, троянські програми для встановлення цільового хоста. – Містить генератор контрольної суми. – Дуже легкий, але потужний інструмент атаки на IRC.
WinCap and JpCap [13]	TCP, UDP, ICMP	TCP dump, UDP ICMP dump	– Програма на базі Windows для передачі мережевого трафіку і процесу стека протоколів

Більшість з перелічених програм доступні безкоштовно. Серед популярних Ddosflowgen та LOIC які виконують DDoS-атаку та генерують синтетичний набори даних трафіку. Середовище OMNET++ дає змогу змоделювати власну топологію мережі. Змодельована мережа дає більше можливостей для вивчення специфіки атак, їх параметрі для формування майбутнього набору даних чи перевірки його.

Висновки

Розглянуто підходи щодо перевірки запропонованих методі виявлення атак. Проаналізовано наявні набори даних, які вчені використовують для створення систем виявлення DDoS-атак. Наведено декілька інструментів, що використовуються для реалізації чи моделювання DDoS-атак для збору даних. Можна використати вже доступний набір даних, але не завжди є необхідна кількість даних чи параметрів. З іншого боку використання програмних засобів для моделювання мережі дає змогу провести будь яку атаку, але дані які отримують з них не наближені до реальних. Результатами моделювання легко аналізувати та корегувати, що дає можливість розробнику гнучкість при формуванні набору даних.

Таким чином, залежно від вибору способу, яким дослідник може отримати набори даних DDoS-атак буде залежати ефективність розробленої IDS проти DDoS-атак.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Voytovych O. P. Denial-of- Service attacks investigation / Voytovych O. P., Kolibabchuk E. I., Kupershtein L. M. // Вісник ХНУ : серія Технічні науки. - №3. -2016. - С. 129-133.

2. Ali, K.. Algorizmi: A Configurable Virtual Testbed to Generate Datasets for O ine Evaluation of Intrusion Detection Systems. Ph.D. thesis – 2010.

3. Ozgur, A. and Erdem, H. A Review of KDD99 Dataset Usage in Intrusion Detection and Machine Learning between 2010 and 2015 – 2016.

4. The CAIDA UCSD "DDoS Attack 2007" Dataset [Назва з екрану]. – Режим доступу до джерела: http://www.caida.org/data/passive/ddos-20070804_dataset.xml

5. Dhanabal, L. and Shantharajah, S.P. A Study of NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms. International Journal of Advanced Research in Computer and Communication Engineering, 4, 446-452. – 2015.

6. Shiravi, A., Shiravi, H., Tavallaee, M. and Ghorbani, A.A. Toward Developing a Systematic Approach to Generate Benchmark Datasets for Intrusion Detection. Computers & Security, 2012, 357-374.

7. University of Southern California-Information Sciences Institute. DoS_80-20110715 (07/15/2011 to 07/15/2011) – 2012.

8. Bhuyan, Monowar H. et al. "Towards Generating Real-life Datasets for Network Intrusion Detection." I. J. Network Security 17: 683-701. – 2015.

9. Войтович О.П., Остапенко-Боженова А.В., Кульчицький Б.В. Виявлення DDoS-атак на основі нейронних мереж. Збірник тез VIII МНТК «Оптоелектронні інформаційні технології «Фотоніка ОДС - 2018», м. Вінниця, 2-4 жовтня 2018 р., С. 179-180.

10. Berkes, J. (2017) Simulating DDoS Attacks with Ddosflowgen. Network Security [назва з екрану]. – Режим доступу до джерела: https://galois.com/blog/2017/04/simulating-ddos-attacks-ddosflowgen/

11. Jonsson, V. (2009) HttpTools: A Toolkit for Simulation of Web Hosts in OMNeT++. Proceedings of the Second International ICST Conference on Simulation Tools and Techniques , Rome, 2 March 2009 [назва з екрану]. – Режим доступу до джерела: https://doi.org/10.4108/ICST.SIMUTOOLS2009.5589

12. Low Orbit Ion Cannon (LOIC) [назва з екрану]. – Режим доступу до джерела: https://github.com/NewEraCracker/LOIC

13. Shinde, P. and Parvat, T.J. (2016) DDoS Attack Analyzer: Using JPCAP and Win-Cap. Procedia Computer Science, 79, 781-784 [назва з екрану]. – Режим доступу до джерела: https://doi.org/10.1016/j.procs.2016.03.103

14. Методичні вказівки до проведення практичних занять та до виконання самостійної й індивідуальної роботи з дисципліни „Основи науково-дослідної роботи/ Укладачі: А. О. Азарова, В. В. Карпінець. – Вінниця: ВНТУ, 2013. – 44 с.

Кульчицький Богдан Володимирович – студент групи БС-18м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, Україна

Куперштейн Леонід Михайлович – к. т. н., доцент кафедри захисту інформації, Вінницький національний технічний університет, м. Вінниця, Україна

Kulchytskyi Bogdan V. – Student of Information Technologies and Computer Engineering epartment, Vinnytsia National Technical University, Vinnytsia, Ukraine

Kupershtein Leonid M. – PhD., Assoc. Professor of Information Protection Chair, Vinnytsia National Technical University, Vinnytsia, Ukraine